TOP PAGE | QUERY | RESULTS | PROJECTS | VIEWS | DATABANKS   **HELP**

Reset

View | * Complete entries * |

This entry is from: DDBJRELEASE:AE000154

DDBJRELEASE

Save

Link

Launch
NClustalW

Printer Friendly

LOCUS	AE000154 11519 bp DNA linear BCT 01-DEC-2000
DEFINITION	Escherichia coli K12 MG1655 section 44 of 400 of the complete genome.
ACCESSION	AE000154 U00096
VERSION	AE000154.1
KEYWORDS	.
SOURCE	Escherichia coli K12
ORGANISM	Escherichia coli K12
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE	1 (bases 1 to 11519)
AUTHORS	Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE	The complete genome sequence of Escherichia coli K-12
JOURNAL	Science 277 (5331), 1453-1474 (1997)
MEDLINE	97426617
PUBMED	9278503
REFERENCE	2 (bases 1 to 11519)
AUTHORS	Blattner,F.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	3 (bases 1 to 11519)
AUTHORS	Blattner,F.R.
TITLE	Direct Submission
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	4 (bases 1 to 11519)
AUTHORS	Plunkett,G. III.
TITLE	Direct Submission
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers

beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES	Location/Qualifiers
source	1..11519
	/organism="Escherichia coli K12"
	/strain="K12"
	/sub_strain="MG1655"
	/db_xref="taxon:83333"
promoter	<1..26
	/note="factor Sigma70; predicted +1 start at 499276"
promoter	complement(2..30)
	/note="factor Sigma70; predicted +1 start at 499238"
gene	106..1410
	/gene="gsk"
	/note="synonym: b0477"
CDS	106..1410
	/gene="gsk"
	/EC_number="2.7.1.73"
	/function="enzyme; Salvage of nucleosides and nucleotides"
	/note="o434; 100 pct identical to GB: ECOGSK_1
	ACCESSION:D00798; 98 pct identical (1 gap) to INGK_ECOLI
	SW: P22937"
	/codon_start=1
	/transl_table=11
	/product="inosine-guanosine kinase"
	/protein_id="AAC73579.1"
	/db_xref="GI:1786684"
	/translation="MKFPGKRKSKHYFPVNARDPLLQQFQPENETSAAWVVGIDQTLV
	DIEAKVDDEFIERYGLSAGHSLVIEDDVAEALYQELKQKNLITHQFAGGTIGNTMHNY
	SVLADDRSVLLGVMCSNIEIGSYAYRYLCNTSSRTDLNYLQGVDGPIGRCFTLIGESG
	ERTFAISPGHMNQLRAESIPEDVIAGASALVLTSYLVRCKPGEPMPEATMKAIEYAKK
	YNVPVVLTLGTKFVIAENPQWWQQFLKDHVSILAMNEDEAEALTGESDPLLASDKALD
	WVDLVLCTAGPIGLYMAGFTEDEAKRKTQHPLLPGAIAEFNQYEFSRAMRHKDCQNPL
	RVYSHIAPYMGGPEKIMNTNGAGDGALAALLHDITANSYHRSNVPNSSKHKFTWLTYS
	SLAQVCKYANRVSYQVLNQHSPRLTRGLPEREDSLEESYWDR"
repeat_region	1425..1508
	/note="REP (repetitive extragenic palindromic) element;
	contains 2 REP sequences"
gene	complement(1543..3219)
	/gene="ybaL"
	/note="synonym: b0478"
CDS	complement(1543..3219)
	/gene="ybaL"
	/function="putative transport; Not classified"
	/note="f558; 100 pct identical to fragment YBAL_ECOLI
	SW:P39830 (70 aa) but has 6 additional C-terminal aa and
	485 additional N-terminal residues"
	/codon_start=1
	/transl_table=11
	/product="putative transport protein"
	/protein_id="AAC73580.1"
	/db_xref="GI:1786685"
	/translation="MHHATPLITTIVGGLVLAFILGMLANKLRISPLVGYLLAGVLAG
	PFTPGFVADTKLAPELAELGVILLMFGVGLHFSLKDLMAVKAIAIPGAIAQIAVATLL
	GMALSAVLGWSLMTGIVFGLCLSTASTVVLLRALEERQLIDSQRGQIAIGWLIVEDLV
	MVLTLVLLPAVAGMMEQGDVGFATLAVDMGITIGKVIAFIAIMMLVGRRLVPWIMARS
	AATGSRELFTLSVLALALGVAFGAVELFDVSFALGAFAGMVLNESELSHRAAHDTLP
	LRDAFAVLFFVSVGMLFDPLILIQQPLAVLATLAIILFGKSLAAFFLVRLFGHSQRTA
	LTIAASLAQIGEFAFILAGLGMALNLLPQAGQNLVLAGAILSIMLNPVLFALLEKYLA
	KTETLEEQTLEEAIEEEKQIPVDICNHALLVGYGRVGSLLGEKLLASDIPLVVIETSR
	TRVDELRERGVRAVLGNAANEEIMQLAHLECAKWLILTIPNGYEAGEIVASARAKNPD
	IEIIARAHYDDEVAYITERGANQVVMGEREIARTMLELLETPPAGEVVTG"
promoter	complement(3328..3356)
	/note="factor Sigma70; predicted +1 start at 502564"
gene	complement(3457..4677)
	/gene="fsr"
	/note="synonym: b0479"
CDS	complement(3457..4677)
	/gene="fsr"

```
                            /function="putative transport; Not classified"
                            /note="f406; 100 pct identical to FSR_ECOLI SW: P52067"
                            /codon_start=1
                            /transl_table=11
                            /product="fosmidomycin resistance protein"
                            /protein_id="AAC73581.1"
                            /db_xref="GI:1786686"
                            /translation="MAMSEQPQPVAGAAASTTKARTSFGILGAISLSHLLNDMIQSLI
                            LAIYPLLQSEFSLTFMQIGMITLTFQLASSLLQPVVGYWTDKYPMPWSLPIGMCFTLS
                            GLVLLALAGSFGAVLLAAALVGTGSSVFHPESSRVARMASGGRHGLAQSIFQVGGNFG
                            SSLGPLLAAVIIAPYGKGNVAWFVLAALLAIVVLAQISRWYSAQHRMNKGKPKATIIN
                            PLPRNKVVLAVSILLILIFSKYFYMASISSYYTFYLMQKFGLSIQNAQLHLFAFLFAV
                            AAGTVIGGPVGDKIGRKYVIWGSILGVAPFTLILPYASLHWTGVLTVIIGFILASAFS
                            AILVYAQELLPGRIGMVSGLFFGFAFGMGGLGAAVLGLIADHTSIELVYKICAFLPLL
                            GMLTIFLPDNRHKD"
promoter                    complement(4711..4738)
                            /note="factor Sigma70; predicted +1 start at 503947"
promoter                    complement(4742..4770)
                            /note="factor Sigma70; predicted +1 start at 503978"
promoter                    complement(4770..4797)
                            /note="factor Sigma70; predicted +1 start at 504006"
promoter                    4833..4861
                            /note="factor Sigma70; predicted +1 start at 504111"
promoter                    complement(4839..4865)
                            /note="factor Sigma70; predicted +1 start at 504075"
gene                        4895..6547
                            /gene="ushA"
                            /note="synonym: b0480"
CDS                         4895..6547
                            /gene="ushA"
                            /EC_number="3.6.1.45"
                            /function="enzyme; Central intermediary metabolism:
                            Sugar-nucleotide biosynthesis, conversions"
                            /note="o550; 99 pct identical to USHA_ECOLI SW: P07024"
                            /codon_start=1
                            /transl_table=11
                            /product="UDP-sugar hydrolase (5'-nucleotidase)"
                            /protein_id="AAC73582.1"
                            /db_xref="GI:1786687"
                            /translation="MKLLQRGVALALLTTFTLASETALAYEQDKTYKITVLHTNDHHG
                            HFWRNEYGEYGLAAQKTLVDGIRKEVAAEGGSVLLLSGGDINTGVPESDLQDAEPDFR
                            GMNLVGYDAMAIGNHEFDNPLTVLRQQEKWAKFPLLSANIYQKSTGERLFKPWALFKR
                            QDLKIAVIGLTTDDTAKIGNPEYFTDIEFRKPADEAKLVIQELQQTEKPDIIIAATHM
                            GHYDNGEHGSNAPGDVEMARALPAGSLAMIVGGHSQDPVCMAAENKKQVDYVPGTPCK
                            PDQQNGIWIVQAHEWGKYVGRADFEFRNGEMKMVNYQLIPVNLKKKVTWEDGKSERVL
                            YTPEIAENQQMISLLSPFQNKGKAQLEVKIGETNGRLEGDRDKVRFVQTNMGRLILAA
                            QMDRTGADFAVMSGGGIRDSIEAGDISYKNVLKVQPFGNVVVYADMTGKEVIDYLTAV
                            AQMKPDSGAYPQFANVSFVAKDGKLNDLKIKGEPVDPAKTYRMATLNFNATGGDGYPR
                            LDNKPGYVNTGFIDAEVLKAYIQKSSPLDVSVYEPKGEVSWQ"
gene                        complement(6584..7063)
                            /gene="ybaK"
                            /note="synonym: b0481"
CDS                         complement(6584..7063)
                            /gene="ybaK"
                            /function="orf; Unknown"
                            /note="f159; 90 pct identical amino acid sequence and
                            equal length to YBAK_SALTY SW: P37174"
                            /codon_start=1
                            /transl_table=11
                            /product="orf, hypothetical protein"
                            /protein_id="AAC73583.1"
                            /db_xref="GI:1786688"
                            /translation="MTPAVKLLEKNKISFQIHTYEHDPAETNFGDEVVKKLGLNPDQV
                            YKTLLVAVNGDMKHLAVAVTPVAGQLDLKKVAKALGAKKVEMADPMVAQRSTGYLVGG
                            ISPLGQKKRLPTIIDAPAQEFATIYVSGGKRGLDIELAAGDLAKILDAKFADIARRD"
promoter                    complement(7211..7239)
                            /note="factor Sigma70; predicted +1 start at 506447"
gene                        complement(7267..8061)
                            /gene="ybaP"
                            /note="synonym: b0482"
CDS                         complement(7267..8061)
```

```
                           /gene="ybaP"
                           /function="putative enzyme; Not classified"
                           /note="f264; This 264 aa ORF is 25 pct identical (10 gaps)
                           to 174 residues of an approx. 976 aa protein SLA2_YEAST
                           SW: P33338"
                           /codon_start=1
                           /transl_table=11
                           /product="putative ligase"
                           /protein_id="AAC73584.1"
                           /db_xref="GI:1786689"
                           /translation="MDLLYRVKTLWAALRGNHYTWPAIDITLPGNRHFHLIGSIHMGS
                           HDMAPLPTRLLKKLKNADALIVEADVSTSDTPFANLPACEALEERISEEQLQNLQHIS
                           QEMGISPSLFSTQPLWQIAMVLQATQAQKLGLRAEYGIDYQLLQAAKQQHKPVIELEG
                           AENQIAMLLQLPDKGLALLDDTLTHWHTNARLLQQMMSWWLNAPPQNNDITLPNTFSQ
                           SLYDVLMHQRNLAWRDKLRAMPPGRYVVAVGALHLYGEGNLPQMLR"
promoter                   7980..8007
                           /note="factor Sigma70; predicted +1 start at 507257"
promoter                   8056..8085
                           /note="factor Sigma70; predicted +1 start at 507335"
gene                       8145..8540
                           /gene="ybaQ"
                           /note="synonym: b0483"
CDS                        8145..8540
                           /gene="ybaQ"
                           /function="orf; Unknown"
                           /note="o131; This 131 aa ORF is 38 pct identical (0 gaps)
                           to 59 residues of an approx. 80 aa protein YSY1_SYNP7 SW:
                           P37371"
                           /codon_start=1
                           /transl_table=11
                           /product="orf, hypothetical protein"
                           /protein_id="AAC73585.1"
                           /db_xref="GI:1786690"
                           /translation="MIQYVLASLFTRKQQLKTMKQATRKPTTPGDILLYEYLEPLDLK
                           INELAELLHVHRNSVSALINNNRKLTTEMAFRLAKVFDTTVDFWLNLQAAVDLWEVEN
                           NMRTQEELGRIETVAEYLARREERAKKVA"
promoter                   complement(8153..8179)
                           /note="factor Sigma70; predicted +1 start at 507389"
repeat_region              8563..8798
                           /note="REP (repetitive extragenic palindromic) element;
                           contains 5 REP sequences"
gene                       complement(8856..11360)
                           /gene="ybaR"
                           /note="synonym: b0484"
CDS                        complement(8856..11360)
                           /gene="ybaR"
                           /function="putative enzyme; Not classified"
                           /note="f834; This 834 aa ORF is 44 pct identical (16 gaps)
                           to 735 residues of an approx. 752 aa protein ATCS_SYNP7
                           SW: P37279"
                           /codon_start=1
                           /transl_table=11
                           /product="putative ATPase"
                           /protein_id="AAC73586.1"
                           /db_xref="GI:1786691"
                           /translation="MSQTIDLTLDGLSCGHCVKRVKESLEQRPDVEQADVSITEAHVT
                           GTASAEQLIETIKQAGYDASVSHPKAKPLAESSIPSEALTAVSEALPAATADDDDSQQ
                           LLLSGMSCASCVTRVQNALQSVPGVTQARVNLAERTALVMGSASPQDLVQAVEKAGYG
                           AEAIEDDAKRRERQQETAVATMKRFRWQAIVALAVGIPVMVWGMIGDNMMVTADNRSL
                           WLVIGLITLAVMVFAGGHFYRSAWKSLLNGAATMDTLVALGTGVAWLYSMSVNLWPQW
                           FPMEARHLYYEASAMIIGLINLGHMLEARARQRSSKALEKLLDLTPPTARLVTDEGEK
                           SVPLAEVQPGMLLRLTTGDRVPVDGEITQGEAWLDEAMLTGEPIPQQKGEGDSVHAGT
                           VVQDGSVLFRASAVGSHTTLSRIIRMVRQAQSSKPEIGQLADKISAVFVPVVVVIALV
                           SAAIWYFFGPAPQIVYTLVIATTVLIIACPCALGLATPMSIISGVGRAAEFGVLVRDA
                           DALQRASTLDTVVFDKTGTLTEGKPQVVAVKTFADVDEAQALRLAAALEQGSSHPLAR
                           AILDKAGDMQLPQVNGFRTLRGLGVSGEAEGHALLLGNQALLNEQQVGTKAIEAEITA
                           QASQGATPVLLAVDGKAVALLAVRDPLRSDSVAALQRLHKAGYRLVMLTGDNPTTANA
                           IAKEAGIDEVIAGVLPDGKAEAIKHLQSEGRQVAMVGDGINDAPALAQADVGIAMGGG
                           SDVAIETAAITLMRHSLMGVADALAISRATLHNMKQNLLGAFIYNSIGIPVAAGILWP
                           FTGTLLNPVVAGAAMALSSITVVSNANRLLRFKPKE"
promoter                   complement(11401..11431)
```

```
                              /note="factor Sigma70; predicted +1 start at 510637"
          promoter            complement(11438..11466)
                              /note="factor Sigma70; predicted +1 start at 510674"
          promoter            11466..11495
                              /note="factor Sigma70; predicted +1 start at 510745"
          promoter            11504..>11519
                              /note="factor Sigma70; predicted +1 start at 510782"
BASE COUNT        2883 a    3112 c    2969 g    2555 t
ORIGIN
        1 gaaaatgcga tcccgcctgc tgatattgaa actggctgcg tctcgcgcgc tcccgtcaga
       61 ttgtgttaac attcgccgct cagttaacca cccgtaaaaa caaccatgaa atttcccggt
      121 aaacgtaaat ccaaacatta cttccccgta aacgcacgcg atccgctgct tcagcaattc
      181 cagccagaaa acgaaaccag cgctgcctgg gtagtgggta tcgatcaaac gctggtcgat
      241 attgaagcga aagtggatga tgaatttatt gagcgttatg gattaagcgc cgggcattca
      301 ctggtgattg aggatgatgt agccgaagcg ctttatcagg aactaaaaca gaaaaacctg
      361 attacccatc agtttgcggg tggcaccatt ggtaacacca tgcacaacta ctcggtgctc
      421 gcggacgacc gttcggtgct gctgggcgtc atgtgcagca atattgaaat tggcagttat
      481 gcctatcgtt acctgtgtaa cacttccagc cgtaccgatc ttaactatct acaaggcgtg
      541 gatggcccga ttggtcgttg ctttacgctg attggcgagt ccggggaacg tacctttgct
      601 atcagtccag gccacatgaa ccagctgcgg gctgaaagca ttccggaaga tgtgattgcc
      661 ggagcctcgg cactggttct cacctcatat ctggtgcgtt gcaagccggg tgaacccatg
      721 ccggaagcaa ccatgaaagc cattgagtac gcgaagaaat ataacgtacc ggtggtgctg
      781 acgctgggca ccaagtttgt cattgccgag aatccgcagt ggtggcagca attcctcaaa
      841 gatcacgtct ctatccttgc gatgaacgaa gatgaagccg aagcgttgac cggagaaagc
      901 gatccgttgt tggcatctga caaggcgctg gactgggtag atctggtgct gtgcaccgcc
      961 gggccaatcg gcttgtatat ggcgggcttt accgaagacg aagcgaaacg taaaacccag
     1021 catccgctgc tgccgggcgc tatagcggaa ttcaaccagt atgagtttag ccgcgccatg
     1081 cgccacaagg attgccagaa tccgctgcgt gtatattcgc acattgcgcc gtacatgggc
     1141 gggccggaaa aaatcatgaa cactaatgga gcgggggatg gcgcattggc agcgttgctg
     1201 catgacatta ccgccaacag ctaccatcgt agcaacgtac caaactccag caaacataaa
     1261 ttcacctggt taacttattc atcgttagcg caggtgtgta aatatgctaa ccgtgtgagc
     1321 tatcaggtac tgaaccagca ttcacctcgt ttaacgccgt gcttgccgga gcgtgaagac
     1381 agcctggaag agtcttactg ggatcgttaa gttatcgtcg gttcgtaggc cagataaggc
     1441 gttcacgccg catctggcat ttggctctcg atgcctgatg cgacgctggc gcgtcttatc
     1501 atgcctacat atttttcata ttttacatcc ggcaaccacc gtttaccccg tcaccacctc
     1561 acccgccggt ggcgtttcca gcagttccag catggtacgg gcgatttcac gctcgcccat
     1621 cactacctga ttcgcaccac gttcggtgat atacgccact tcatcgtcat aatgggcgcg
     1681 ggcaataatc tcaatatccg gattttttcgc gcgggcagat gccacaatct cacccgcttc
     1741 ataaccgttg ggaatcgtca ggatcagcca ttttgcacat tccagatgcg ccagttgcat
     1801 aatttcttcg ttcgccgcat tgcccaatac tgcgcggacc ccgcgctctc gcagctcatc
     1861 aacacgggtt cgtgacgtct caatcaccac cagcggaata tcagaggcga gcaatttctc
     1921 ccccagcagg ctgcctacac gaccgtaacc caccagtagc gcatggttgc aaatatccac
     1981 tgggatctgc ttctcttctt cgattgcctc ttccagcgtc tgctcttcca gcgtttcggt
     2041 cttcgccaga tatttctcca gtagtgcgaa cagtaccggg ttgagcataa tcgacaggat
     2101 cgcccctgcc agtaccaggt tttgtccggc ctgcggcagt aaattcaatg ccattcccag
     2161 tcccgccagg ataaacgcga actcaccaat ctgcgccagg ctggcggcga tggttaatgc
     2221 cgtacgttgg gagtgaccaa acagtcgcac caggaaaaat gcggctaacg acttaccaaa
     2281 cagaataatc gccagcgtcg ccagcactgc cagcggttgc tgaatcagaa ttaacggatc
     2341 aaacaacatc ccgacggaga caaaaaacag caccgcaaac gcgtcgcgca atggcagcgt
     2401 atcgtgggcg gcacggtgac tcagttcaga ctcgttcagt accatcccgg caaagaacgc
     2461 accgagtgca aaggagacat caaacagctc taccgcacca aaggcaaccc ctaacgccag
     2521 cgccagcacc gacagggtaa acagctcgcg agaaccggtt gccgcgctgc gtgccataat
     2581 ccacgcacc agacggcgac ctaccagcat cataatggcg ataaatgcga tcactttgcc
     2641 gatggtgatc cccatatcga ctgcaagagt ggcaaagccc acatcgccct gttccatcat
     2701 tcctgccact gcgggcagca acaccagcgt cagaaccatt accaggtctt ccacaatcaa
     2761 ccaaccgatg gcgatttgcc cacgctgact gtcaattaat tgccgttctt caagtgcgcg
     2821 cagtaacacc acggtactgg cggtggaaag acataaaccg aacacgatac cggtcattaa
     2881 cgaccagccc agcacggcag agagcgccat acccagcagc gtcgccacgg ctatctgggc
     2941 gatcgcaccg ggaatggcga tggcctttac cgccatcaaa tccttcagcg aaaagtgcaa
     3001 accgacgcca aacatcaaca gaatgacgcc cagttcagcc agttccgggg caagcttggt
     3061 atcggcaaca aagcccggag tgaatggtcc tgccagcaca cccgctaaca gatatcccac
     3121 cagaggagaa atacgtagtt tattggccag catgccgagg ataaaggcga gcacaaggcc
     3181 gccaacaatg gtggtgataa gcggggtggc gtgatgcatt ccgtctcctt ttcctggtgg
     3241 ttattgtcca tttttggccg ggaaaaccaa aattacaggt aatagtttat gacaatttca
     3301 ttgatgatgt tcatgaataa ttattgaatt ttgcagaaaa atggaattag ctgcaaaaaa
     3361 agcacggatc ggaaaacaga aggcggttta acagaggaaa ggcttatggc gactgtaaaa
     3421 cgatgcagcc aaagtttgcc tttggctgca atgaaatcag tctttatgcc ggttatcagg
     3481 caggaatatg gtcaacatcc ccaatagtgg caggaaagca cagattttat agactaactc
     3541 gatgctggtg tgatcggcga taagcccag aactgccgct cccagacctc ccatgccaaa
     3601 agcaaaaccg aaaaagagtc cagaaaccat accgatacgt cctggaagca gctcctgagc
     3661 gtagaccaga atggcagaga atgccgaagc gaggataaat ccaataatca ccgttaaaac
```

```
3721  ccccgtccag tgcaggctgg cgtagggtaa aatcagcgta aacggcgcaa cgccgaggat
3781  agagccccaa atcacatatt tccgcccaat tttatcccct acaggcccgc cgatcaccgt
3841  acctgccgca acggcaaaca ggaaggcaaa cagatgaagc tgagcattct ggatagataa
3901  tccgaatttt tgcatcagat aaaaggtgta atagctgctg atgctcgcca tatagaaata
3961  tttcgagaaa atgaggatta acagaatgct gaccgccagt acaactttat tgcgcggcag
4021  tggattgata atcgtcgctt tgggttttcc tttattcatt cggtgctgtg ccgagtacca
4081  acggctgatt tgcgccaaca ccacgatcgc cagcagtgcc gcaagcacaa accaggcaac
4141  gttgcctttg ccataaggcg cgataatcac cgccgccagc aagggtccca gggaactgcc
4201  aaagttgccg ccgacctgaa agatagattg cgccaggcca tgccgcccgc cggaagccat
4261  acgggccacg cgagaagatt ccggatgaaa gaccgatgaa ccggtaccga ccagcgccgc
4321  cgccagcaga actgcgccaa aactgcccgc cagcgcaagc agcaccagac cgcttaaggt
4381  aaagcacatg ccaattggca acgaccatgg catcggatat ttatcggtcc agtagccgac
4441  cactggttgc agtagcgaag aggcgagctg gaaggtgagg gttatcatgc caatctgcat
4501  aaatgtcaga gaaaattctg actgaagcag cggataaatc gccagaatca gcgattggat
4561  catgtcgttc agcagatgtg agaggctgat agcacctaaa ataccaaacg atgttcgggc
4621  cttggtcgtt gacgcagccg cgcccgccac aggctggggt tgttcactca ttgccatagg
4681  aaagtcactt tttcagggtt gcgatgtaaa gaatgatctt atttgtgatt attaccagac
4741  taacatacct gtatgcgtcg tctgaaggaa gtctcaacgc cgaatacaga atttctaatc
4801  tggatgcaga tttatcttca ccggacgcag acttgtctat gatgtcgcgt catactattt
4861  ttcaacacgt tgaaatcagg tcagggagag aagtatgaaa ttattgcagc ggggcgtggc
4921  gttagcgctg ttaaccacat ttacactggc gagtgaaact gctctggcgt atgagcagga
4981  taaaacctac aaaattacag ttctgcatac caatgatcat catgggcatt tttggcgcaa
5041  tgaatatggc gaatatggtc tggcggcgca aaaaacgctg gtggatggta tccgcaaaga
5101  ggttgcggct gaaggcggta gcgtgctgct actttccggt ggcgacatta acactggcgt
5161  gcccgagtct gacttacagg atgccgaacc tgattttcgc ggtatgaatc tggtgggcta
5221  tgacgcgatg gcgatcggta atcatgaatt tgataatccg ctcaccgtat tacgccagca
5281  ggaaaagtgg gccaagttcc cgttgctttc cgcgaatatc taccagaaaa gtactggcga
5341  gcgcctgttt aaaccgtggg cgctgtttaa gcgtcaggat ctgaaaattg ccgttattgg
5401  gctgacaacc gatgacacag caaaaattgg taacccggaa tacttcactg atatcgaatt
5461  tcgtaagccc gccgatgaag cgaagctggt gattcaggag ctgcaacaga cagaaaagcc
5521  agacattatt atcgcggcga cccatatggg gcattacgat aatggtgagc acggctctaa
5581  cgcaccgggc gatgtggaga tggcacgcgc gctgcctgcc ggatcgctgg cgatgatcgt
5641  cggtggtcac tcgcaagatc cggtctgcat ggcggcagaa aacaaaaaac aggtcgatta
5701  cgtgccgggt acgccatgca aaccagatca acaaaacggc atctggattg tgcaggcgca
5761  tgagtggggc aaatacgtgg gacgggctga ttttgagttt cgtaatggcg aaatgaaaat
5821  ggttaactac cagctgattc cggtgaacct gaagaagaaa gtgacctggg aagacgggaa
5881  aagcgagcgc gtgctttaca ctcctgaaat cgctgaaaac cagcaaatga tctcgctgtt
5941  atcaccgttc cagaacaaag gcaaagcgca gctggaagtg aaaataggcg aaaccaatgg
6001  tcgtctggaa ggcgatcgtg acaaagtgcg ttttgtacag accaatatgg ggcggttgat
6061  tctggcagcc caaatggatc gcactggtgc cgactttgcg gtgatgagcg gaggcggaat
6121  tcgtgattct atcgaagcag gcgatatcga ctataaaaac gtgctgaaag tgcagccatt
6181  cggcaatgtg gtggtgtatg ccgacatgac cggtaaagag gtgattgatt acctgaccgc
6241  cgtcgcgcag atgaagccag attcaggtgc ctacccgcaa tttgccaacg ttagctttgt
6301  ggcgaaagac ggcaaactga acgaccttaa aatcaaaggc gaaccggtcg atccggcgaa
6361  aacttaccgt atggcgacat taaacttcaa tgccaccggc ggtgatggat atccgcgcct
6421  tgataacaaa ccgggctatg tgaataccgg ctttattgat gccgaagtgc tgaaagcgta
6481  tatccagaaa agctcgccgc tggatgtgag tgtttatgaa ccgaaaggtg aggtgagctg
6541  gcagtaatcc gaaagtgccg gatgtttgca tccggcacaa tgcttaatcg cggcgggcga
6601  tatcagcaaa tttggcatcg aggatctttg ccagatcgcc tgccgccagt tcgatatcca
6661  gtccgcgctt gccgccggaa acataaatag tggcaaattc ttgtgcgggg gcgtcgataa
6721  tcgttggcag acgttttttc tgccccagtg gctaatccc cccaaccagg tatcccgtcg
6781  aacgctgcgc gaccatcgga tcggccatct caactttctt ggcacccagc gcttttgcta
6841  ctttttaag atccagttga cctgcgaccg gcgtaacggc cacggcaagg tgtttcatat
6901  caccgttcac tgccaccagc agcgttttgt agacctgatc cggattcaaa cctaattttt
6961  tgacgacttc atcgccaaaa ttggtttcag ccggatcgtg ctcgtaggta tggatttgaa
7021  acgaaatctt gtttttttcg agtaatttaa ctgcgggtgt catggtaatt ttcctgcctt
7081  aaacaaaaaa acgaagcaag catacgccta atggcggtcc aaaaaaaaga gtcatcttgc
7141  ctaagagtat tggcaggatg gtgagattga gcgacaatcg agttacaccg tcgcttaaag
7201  tgacggcata ataataaaaa aatgaaattc ctctttgacg ggccaatagc gatattggcc
7261  attttttag cgcaacattt gcggcaaatt cccttctcca tacaggtgta gtgcaccgac
7321  cgcgaccaca tatcgcccog gcggcatggc gcgtaattta tcccgccagg cgagatttcg
7381  ctgatgcatc agcacatcgt acagcgactg actgaacgta ttgggcagcg ttatatcatt
7441  attttgcggc ggtgcattca gccaccagct catcatttgt tgcagcaacc gtgcgttggt
7501  atgccagtgg gtcagcgtat cgtccagcag cgccagtcct ttgtcaggga gctggagcaa
7561  catggcaatc tggttttcag cccccttccag ttcaatcacg ggtttatgtt gttgcttcgc
7621  cgcctgcaat agctggtaat cgataccgta ttctgcccgc agccccagtt tttgtgcctg
7681  cgtcgcctga agaaccatcg cgatttgcca cagcggttgg gtagaaaaga gtgagggaga
7741  aatgcccatc tcctggctaa tgtgctgcag gtttttggagt tgctcctcgc taatgcgctc
7801  ttccagcgcc tcgcaggcag gcaaattagc aaaaggcgta tcgctggtgg aaacatccgc
7861  ctcgacgatc agcgcatcgg cgtttttgag ctttttgagc aaacgggtgg gcaggggagc
7921  catatcgtgg ctacccatat gaatactgcc aatcagatga aaatggcgat tgccgggggag
```

```
 7981  ggtgatatcg atggcaggcc aggtgtaatg attaccgcgc agcgcggccc aaagcgtttt
 8041  tacccggtac aacagatcca tacgacctcc ctttgtgaaa tatcatgcta gcgcgcggtg
 8101  agggatggcg caaatggagt tggagacttg tttaatgtgt ttgtatgatt cagtatgttc
 8161  ttgcatcgct attcacaagg aagcaacagt taaaaaccat gaaacaggca acaagaaaac
 8221  cgacgacacc cggagatatt cttctctatg aatatctgga accgctcgat ttgaaaatca
 8281  atgagttagc agagttgctg catgttcatc gtaatagcgt cagtgcactg atcaataaca
 8341  atcgtaaact cactactgag atggcatttc gtctggcgaa agtttttgat accacagtcg
 8401  atttttggct aaacctccag gcggcggttg atctttggga agttgaaaac aacatgcgca
 8461  cccaggaaga attgggacgg attgaaacag tggctgaata tttggcacgc cgtgaagagc
 8521  gtgcaaaaaa ggtcgcgtaa ggccaaaaca tcgcttcatc tcgtaggccg gataaggcgt
 8581  tcacgccgca tccggcatcc gagcaccatt gcctgatgcg acgcttaccc gtcttatcag
 8641  gcctacaaac ttgtgcctga accgtaggcc ggataaggcg tttacgccgc atccggcatc
 8701  cgagcaccat tgcctgatgc gacgcttacg cgtcttatca ggcctacaaa cctgtgcctg
 8761  aaccgtaggc cggataaggc gtttacgccg catccggcag tcatgcgtcg atgccaaatg
 8821  cgccaccta aagcagcgca tccgcaatga tgtacttatt ccttcggttt aaaccgcagc
 8881  aaccggttgg cgttactcac tacggtaatc gacgagagcg ccattgccgc tccggcaact
 8941  accgggttaa gcagtgttcc agtgaacggc cacaaaatac cggcggcgac cggaataccg
 9001  atactgttgt agataaacgc accgagcagg ttctgcttca tgttgtgcag cgttgcgcgg
 9061  gaaatagcga gcgcatccgc aacgcccatc aggctatggc gcatcagggt aatcgccgcg
 9121  gtttcaatgg caacatcact gccgccaccc atcgcaatgc cgacatccgc ctgagccagc
 9181  gctggcgcgt cgttaatgcc gtcgcccacc attgccacct gacgtccttc actttgcaga
 9241  tgtttgatcg cttcggcttt accgtccggc agcaccccgg cgatcacctc atcaatccct
 9301  gcttcttgg cgatcgcatt ggcggtggtt gggttatccc cggtcaacat caccagacga
 9361  tatcccgctt tatgcaggcg ttgcagcgcc gccacgctat cactacgcaa cggatcgcgt
 9421  actgccagca gggctaccgc tttcccgtca accgccagca gcacaggcgt tgcccctgc
 9481  gatgcctgag cagtaatctc cgcttcgata gctttggtac caacctgttg ctcatttaac
 9541  agcgcctgat tgcccagcaa taacgcatga ccttcagctt caccgctcac gcccagcccg
 9601  cgcaatgtgc ggaaaccgtt gacctgcggt agctgcatat cacctgcttt atcgaggatc
 9661  gctcgtgcca gcggatggct ggaaccttgc tccagtgccg ccgccagacg caatgcctgc
 9721  gcttcatcaa catcagcaaa tgtttttcact gcgacaacct gcggcttccc ttcagtcagc
 9781  gtcccggttt tatcgaacac tacagtgtcg agtgtactgg cgcgttgcag cgcgtcagcg
 9841  tcccgcacca gcacgccaaa ctcagccgcc cgcccgacgc cggaaataat cgacatcggc
 9901  gtcgccagcc ccagcgcaca cggacaggca ataatcagta ccgtggtggc aatcaccagg
 9961  gtatagacaa tctgcggtgc cggaccaaag aaataccaga ttgccgcact gacaagcgca
10021  ataaccacca ctaccggcac aaatacggct gagattttat ccgccagctg accgatttct
10081  ggcttgctgc tctgggcctg gcgcaccatg cgaatgattc gtgacagcgt agtatggctg
10141  ccaaccgcac tggcacgaaa cagcacactg ccgtcctgta ccactgtccc ggcatggacg
10201  ctatcgcctt cgccttttttg ctgcgggatt ggttcgcccg tcagcatcgc ttcatccagc
10261  catgcttcgc cctgggtaat ctcgccatct accggcacgc gatcgccggt cgtcaggcgc
10321  agcaacatac ctggctgcac ttctgccaga ggcacgcttt tttcaccttc gtcagtaacc
10381  aggcgtgccg tcggcggggt taaatcgagt aacttttcca gcgccttaga agaacgctgg
10441  cgtgcgcgcg cttccagcat atggccgaga ttgatcagac cgataatcat cgcgctggct
10501  tcgtaataaa gatgtcgcgc ttccatcggg aaccactgcg gccacaggtt gacgctcatc
10561  gaatagagcc acgccacgcc agtacccagc gccaccagcg tatccatcgt cgccgcaccg
10621  ttcagcaggc tttttccatgc actgcggtaa aaatggccgc cggcgaaaac catcactgcc
10681  agggttatca ggccgataac caaccacagg ctgcggttgt cagcggtgac catcatgtta
10741  tcgccgatca tcccccagac catcaccggg atacccaccg ccagtgcgac aattgcctgc
10801  cagcggaagc gcttcatcgt agcgacggcg gtttcttgct ggcgctcgcg gcgtttagcg
10861  tcatcttcaa tcgcttccgc gccgtagccc gcttttttcca ccgcctgcac taaatcttgt
10921  ggggaggcac tgcccatcac cagcgcagta cgctccgcca ggtttacccg tgcctgagtg
10981  acgcccggta cgctttgcag cgcatttttgt acgcgggtga cacagctggc gcagctcatg
11041  ccgctcagca gcaactgctg gctgtcatcg tcatcggcgg tcgctgccgg aagcgcctca
11101  gaaaccgctg tcagtgcttc cgacgggatt gatgactccg ccagcggttt agcctttggg
11161  tggcttacag atgcgtcata acccgcttgt ttgatggttt cgatcagctg ttctgcactg
11221  gcagtcccgg taacgtgcgc ttcagtgata gacacatccg cctgctcaac atccggacgc
11281  tgttcaagac tttctttcac gcgtttaacg cagtgaccgc aggacaggcc gtccagggtc
11341  aggtcgatag tttgtgacat aaaacactcc tttaagacag ttttgactgg ctgtgataaa
11401  ggttaaacct tccagcaagg ggaaggtcaa gaaattaata aaccaggcgg gtaaaagtcc
11461  gtaaagatta aaaaatcggc tcgatttgca tcaggattag acatttatct ctttgtttt
//
```